



HAL
open science

Arabidopsis seed content QTL mapping using high-throughput phenotyping: the assets of near infrared spectroscopy

Sophie Jasinski, Alain Lecureuil, Monique Durandet, Patrick Bernard-Moulin,
Philippe Guerche

► To cite this version:

Sophie Jasinski, Alain Lecureuil, Monique Durandet, Patrick Bernard-Moulin, Philippe Guerche. Arabidopsis seed content QTL mapping using high-throughput phenotyping: the assets of near infrared spectroscopy. *Frontiers in Plant Science*, 2016, 7 (1682), pp.1-11. 10.3389/fpls.2016.01682 . hal-02637445

HAL Id: hal-02637445

<https://hal.inrae.fr/hal-02637445>

Submitted on 27 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



Arabidopsis Seed Content QTL Mapping Using High-Throughput Phenotyping: The Assets of Near Infrared Spectroscopy

Sophie Jasinski^{1*}, Alain Lécureuil¹, Monique Durandet¹, Patrick Bernard-Moulin² and Philippe Guerche¹

¹ Institut Jean-Pierre Bourgin, INRA, AgroParisTech, CNRS, Université Paris-Saclay, Versailles, France, ² ThermoFisher Scientific, Courtaboeuf, France

OPEN ACCESS

Edited by:

Marcos Egea-Cortines,
Universidad Politécnica de Cartagena,
Spain

Reviewed by:

Veronique Storme,
Ghent University, Belgium
Yuhui Chen,
Samuel Roberts Noble Foundation,
USA

*Correspondence:

Sophie Jasinski
sophie.jasinski@versailles.inra.fr

Specialty section:

This article was submitted to
Technical Advances in Plant Science,
a section of the journal
Frontiers in Plant Science

Received: 08 September 2016

Accepted: 25 October 2016

Published: 10 November 2016

Citation:

Jasinski S, Lécureuil A, Durandet M,
Bernard-Moulin P and Guerche P
(2016) Arabidopsis Seed Content
QTL Mapping Using High-Throughput
Phenotyping: The Assets of Near
Infrared Spectroscopy.
Front. Plant Sci. 7:1682.
doi: 10.3389/fpls.2016.01682

Seed storage compounds are of crucial importance for human diet, feed and industrial uses. In oleo-proteaginous species like rapeseed, seed oil and protein are the qualitative determinants that conferred economic value to the harvested seed. To date, although the biosynthesis pathways of oil and storage protein are rather well-known, the factors that determine how these types of reserves are partitioned in seeds have to be identified. With the aim of implementing a quantitative genetics approach, requiring phenotyping of 100s of plants, our first objective was to establish near-infrared reflectance spectroscopic (NIRS) predictive equations in order to estimate oil, protein, carbon, and nitrogen content in *Arabidopsis* seed with high-throughput level. Our results demonstrated that NIRS is a powerful non-destructive, high-throughput method to assess the content of these four major components studied in *Arabidopsis* seed. With this tool in hand, we analyzed *Arabidopsis* natural variation for these four components and illustrated that they all displayed a wide range of variation. Finally, NIRS was used in order to map QTL for these four traits using seeds from the *Arabidopsis thaliana* Ct-1 × Col-0 recombinant inbred line population. Some QTL co-localized with QTL previously identified, but others mapped to chromosomal regions never identified so far for such traits. This paper illustrates the usefulness of NIRS predictive equations to perform accurate high-throughput phenotyping of *Arabidopsis* seed content, opening new perspectives in gene identification following QTL mapping and genome wide association studies.

Keywords: *Arabidopsis thaliana*, seed storage contents, near infrared spectroscopy, plant, natural variation, quantitative trait loci

BACKGROUND

Plant seeds constitute a key component of both human and livestock diets, as seed storage compounds are mainly composed of protein, oil and starch. Seed oil from oleaginous crops are composed mainly of triacylglycerols, which are structurally similar to long chain hydrocarbons derived from petroleum, and thus represent ecologically and economically competitive alternatives

Abbreviations: QTL, quantitative trait loci; RIL, recombinant inbred line.

to petroleum-based products for the production of molecules for green chemistry (e.g., detergents, paints, plastics, and lubricants) as well as for the production of biofuels (Durrett et al., 2008; Dyer et al., 2008). The increasing demand of plant-derived products for nutritional and industrial applications highlights the urgent need to develop new methodologies to increase the overall seed oil and protein content. Although, most of the biochemical steps involved in oil and protein biosynthesis are known and the key genes have been identified (Shewry et al., 1995; Beisson et al., 2003), the regulation of the processes that results in the final oil and protein content is not well-understood. Even more, the genetic factors that control the oil/protein ratio in the seeds have to be identified.

Seed oil and protein accumulation processes, like many important agronomical traits are quantitative and have a complex genetic basis. The method most commonly used for inferring the presence and position of such genes in the genome is based upon analysis such as QTL and more recently genome wide association studies (GWASs). Quantitative genetics has been used to search for genetic factors controlling oil and/or protein quantity in a variety of agronomically important species including rapeseed (Sun et al., 2012; Li et al., 2014), soybean (Eskandari et al., 2012; Hwang et al., 2014), maize (Zheng et al., 2008; Li et al., 2012; Yang et al., 2012), pea (Irzykowska and Wolko, 2004; Tar'an et al., 2004), rice (Ying et al., 2012), wheat (Plessis et al., 2013), oat (Kianian et al., 1999), sunflower (Mokrani et al., 2002), linseed (Kumar et al., 2015), cotton (Liu et al., 2015), and *Jatropha* (Liu et al., 2011). However, with the exception of one QTL affecting seed oil content in maize (Zheng et al., 2008), none of these studies have gone beyond the gene mapping stage. In the model species *Arabidopsis*, using natural variation resources, QTL involved in seed oil content and/or quality have also been detected (Hobbs et al., 2004; Jasinski et al., 2012; O'Neill et al., 2012; Sanyal and Linder, 2012) as well as "regions of interest" by GWAS (Branham et al., 2015), but only the study published by Jasinski et al. (2012) identified the gene involved. Since QTL cloning is often easier in model species for which substantial genetic resources exist, we implemented a QTL approach to study storage compound metabolisms in *Arabidopsis* seed (Jasinski et al., 2012; Chardon et al., 2014). Seed metabolism is very similar between *Arabidopsis* and *Brassica* species and the close relationship between them allows the use of comparative genetics to predict orthologous genes and alleles within the *Brassica* genome (Parkin et al., 2005). This will enable the translation of discoveries from *Arabidopsis* into Brassicaceae and other crops breeding programs.

Quantitative genetics relies on statistical links between phenotype and genotype, implying genotyping and phenotyping of 1000s of lines. In *Arabidopsis*, genotyping is not a limiting factor and many genetic and genomic resources are available including complete genome sequence of many accessions, data on gene structure, gene expression, DNA and seed stocks, genome maps, molecular markers. Seed oil and protein content are usually determined by standard analytical methods such as Soxhlet or gas chromatography (following fatty acid methyl ester extraction)

for oil content and combustion analysis or Kjeldahl for protein content. Although, these standard analytical techniques offer a high level of accuracy and precision, they also show some limitations, such as indirect determination (combustion analysis and Kjeldahl determine nitrogen content rather than actual protein content), high costs, time-consuming experiments and use of hazardous chemicals. For many of these reasons, they are not fully appropriate for high-throughput phenotyping required in genetics approaches. Near infrared spectroscopy (NIRS) is a vibrational spectroscopy technique, providing a spectrum representative of the "signature" of all components present in the analyzed sample. It possesses numerous advantages compared to classical analytical techniques. NIRS analyses show high degree of repeatability and are carried out with considerable saving of time (spectrum acquisition lasts only a few seconds), cost and without using hazardous chemicals. In addition, samples can be analyzed in their natural form without destruction neither any special sample preparation. However, a calibration has first to be established: regression modeling is used to relate NIRS spectra to chemical concentrations determined by a standard analytical method. After calibration, the developed regression equations allow accurate analysis of many other samples by prediction of data based on the spectra. Moreover, from only one spectrum, different components can be predicted using different predictive equations. In recent decades, NIRS has been widely used as a fast and reliable method for qualitative and quantitative analysis in many fields (Font et al., 2006) and International Standards Committees have formally accepted methods using NIRS for analysis of many compounds (Batten, 1998). Regarding *Brassica* seeds, many authors have reported NIRS models for different components, such as glucosinolates (Velasco and Becker, 1998; Font et al., 2004), fiber (Font et al., 2003), protein and oil contents (Tkachuk, 1981; Font et al., 2002a,b; Rossato et al., 2013).

Surprisingly, NIRS technique has not been applied to the analysis of *Arabidopsis* seed. Some people used nuclear magnetic resonance spectroscopy (NMR) as rapid technique to measure *Arabidopsis* seed oil content (O'Neill et al., 2003, 2012; Hobbs et al., 2004). However, NMR is not suitable for protein detection and was thus not suitable for our purpose. In this study, the potential of NIRS was evaluated for the simultaneous analysis of total oil and protein content of *Arabidopsis* seeds, as well as nitrogen and carbon contents, which allow studies of global metabolic fluxes. A calibration set of 90-112 seed samples was subjected to both NIRS and appropriate reference methods and predictive equations for seed (1) oil, (2) protein, (3) carbon, and (4) nitrogen content were developed.

These equations were further used to analyze *Arabidopsis* natural variation for these four major seed components. Finally, a search for genetic factors governing the accumulation of these four components in *Arabidopsis* seed was carried out by a QTL analysis (this work and Chardon et al., 2014), allowing the mapping of new QTL involved in seed oil and protein content.

RESULTS

Development of NIRS Predictive Equations for Seed Oil, Protein, Carbon, and Nitrogen Content

Four calibration models in order to predict oil, protein, carbon, and nitrogen content in *Arabidopsis* seeds were developed as indicated in Section “Methods.”

The oil calibration set of 112 samples showed a wide range of variation for oil content from 18.70%, corresponding to the *wri1* low-seed-oil T-DNA insertion mutant (Focks and Benning, 1998), to 46.90%, with a mean of 38.78% (Table 1) and a standard deviation (SD) of 3.88%. The predictive equation for seed oil content was developed with five partial least square (PLS) factors and first evaluated through cross-validation (leave-one-out method). Very high coefficient of determinations between Soxhlet and NIRS values were observed for both calibration and cross-validation ($r_C^2 = r_{CV}^2 = 0.98$, Table 1; Figure 1A). The standard error of cross-validation (SECV) was 0.606% (Table 1). The 36 additional seed samples were used to carry out an external validation to better assess the accuracy of this calibration model. This showed a coefficient of determination of 0.99 and a standard error of prediction (SEP) of 0.505% (Table 1; Figure 1A).

As for oil, the three other calibration sets showed a wide range of variation (Table 1). For each of the three components, a predictive equation was developed and the external validation sample set was further used to evaluate its performance. For the three developed models, a limited number of PLS factors (≤ 4) was used and very high coefficients of determination between the reference method and NIRS values were observed for both calibration and cross-validation (Table 1; Figure 1).

The ratio of performance deviation (RPD), an indicator for the usefulness of the calibration model was calculated for each component (Table 1). According to the American Association of Cereal Chemists Method-39-00.01 (AACCC International, 1999), a RPD ≥ 2.5 indicates that a calibration equation is useful for screening in breeding programs, a RPD ≥ 5 that a calibration is acceptable for quality control and a RPD ≥ 8 that a calibration is good for process control, development, and applied research. For all the models developed in this study, a RPD > 2.5 was achieved, with a RPD of 7.68 (close to 8) for the oil content model and a RPD close to 5 for the nitrogen model. This indicates that the four models developed in this study are suitable for quantitative genetic approaches.

NIRS Is a Suitable Tool for High-Throughput Phenotyping of *Arabidopsis* Seed

In order to fully demonstrate that the developed NIRS models were suitable to study seed composition in *Arabidopsis*, we analyzed mutants altered in seed filling. Pyruvate kinase (PK) catalyze the irreversible synthesis of pyruvate and ATP (Valentini et al., 2000), which are essential for fatty acid production in the plastids of maturing *Arabidopsis* embryos. Baud et al. have shown that the plastidial PK isoform PKp2 plays an important role in

seed oil synthesis, with *pkp2-1* mutant exhibiting a 50% reduction in seed oil content compared to wild-type (Baud et al., 2007). More recently, Chen et al. showed that seed filling in *Arabidopsis* requires sucrose transporters from the SWEET family (Chen et al., 2015). In particular, they showed that seed oil content was reduced by 34% in the *sweet11;12* double mutant (Chen et al., 2015). Seeds from *pkp2-1* and *sweet11;12* mutants were analyzed by NIRS. For *pkp2-1*, a 36% decrease in oil content compared to wild-type was observed (Figure 2A), which is comparable to the decrease described by Baud et al. (2007) on the same seed lot. *Sweet11;12* mutants displayed a 17% reduction in seed oil content compared to wild-type (Figure 2B), which is half the one described by Chen et al. (2015) on another seed lot. The *sweet11;12* seed lot measured by NIRS was then subjected to Gas Chromatography and resulted in a 15% decrease in seed oil (result not shown). This result suggested that the difference observed is probably due to environmental effect on seed filling more than to NIRS method.

Furthermore, since it is known that nitrogen nutrition impacts seed filling (Masclaux-Daubresse and Chardon, 2011), we analyzed seed content of wild-type plants (Col-0 and Ws) grown under low nitrogen (LowN; 2 mM nitrate) or under high nitrogen (HighN; 10 mM nitrate) nutrition conditions. As already published (Masclaux-Daubresse and Chardon, 2011), both accessions displayed higher seed nitrogen content under HighN compared to LowN (Figure 2C) and reversely a higher seed carbon content under LowN compared to HighN (Figure 2D).

These results demonstrate that NIRS is a powerful method to determine *Arabidopsis* seed composition.

Natural Variation for Oil, Protein, Carbon, and Nitrogen Content in *Arabidopsis* Seed

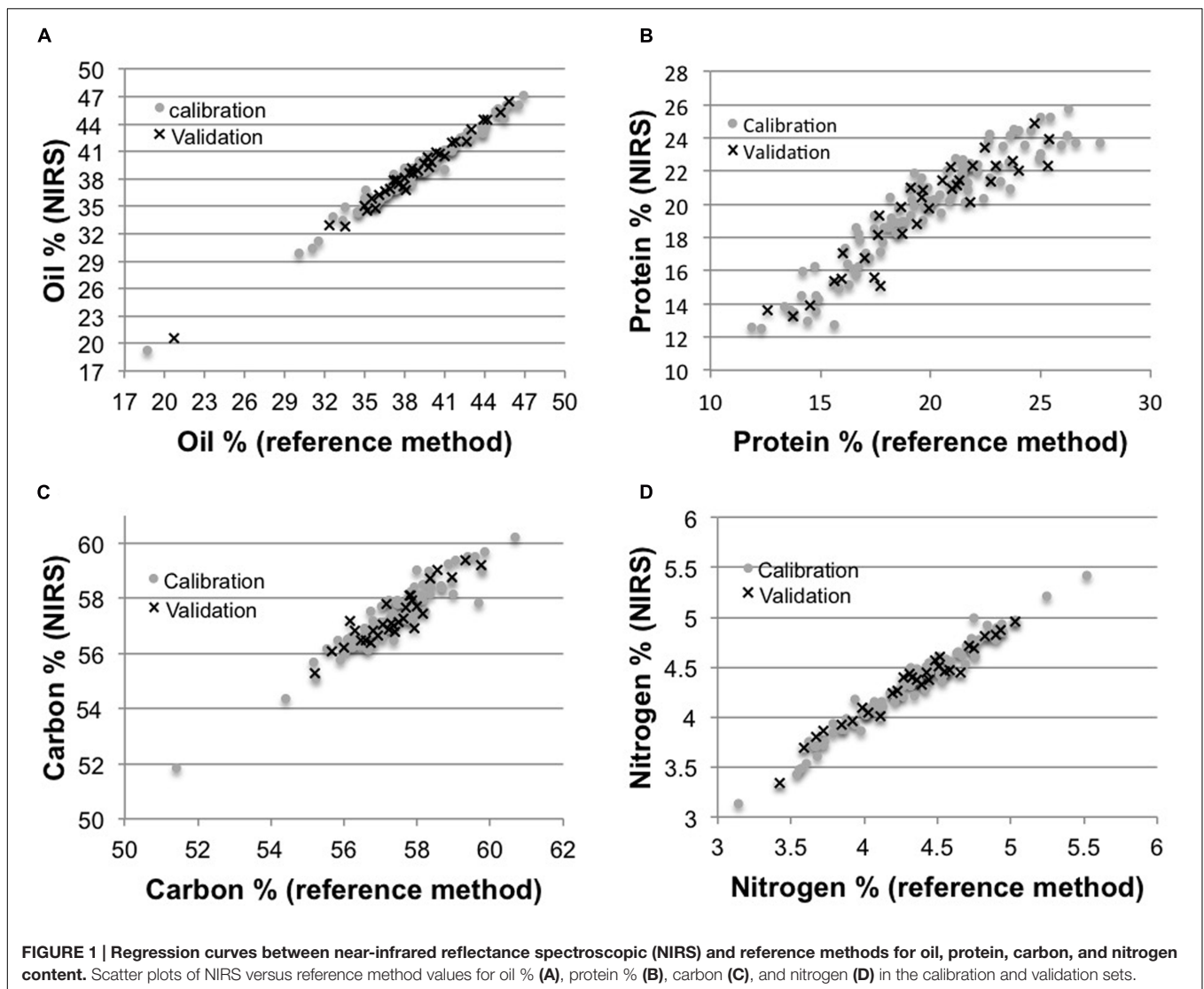
The development of NIRS predictive equations allowing high-throughput phenotyping opened the door to quantitative genetics study. First, we decided to explore *Arabidopsis* natural variability for oil, protein, carbon, and nitrogen content in seeds. For this purpose, we cultivated the Versailles BRC 48 core-collection of *Arabidopsis* in addition to the Col-0 accession and mini sets of 20 lines (maximizing genotypic variability, Simon et al., 2008) from eight populations (see Methods). Each genotype was cultivated in triplicate and three successive and independent cultures (C1, C2, and C3) were performed in growth chambers with similar global climate conditions. We estimated the natural variability of the four traits by NIRS phenotyping (Figure 3). The four traits displayed a wide range of variation in each culture, with C1 displaying the wider range going from 23.23 to 47.72% for oil, from 14.24 to 28.72% for protein, from 52.27 to 60.22% for carbon, and from 3.27 to 5.78% for nitrogen. The modal class is different in each culture, highlighting the environmental effect on these four traits.

In order to quantify the relative contribution of the genotype (G), the culture (C) and the G*C interaction on the variation of these four traits, a global analysis of variance (ANOVA) was performed on the measures from the three cultures (Figure 4). The genotypic effects ranged from 53.0% for oil to 32.3%

TABLE 1 | Near-infrared reflectance spectroscopic (NIRS) calibration and cross validation statistics for seed oil, protein, carbon, and nitrogen contents (%).

	Calibration					Cross-validation				External validation					
	<i>n</i>	Mean (%)	Range (%)	SD	PLS factors	SEC	r_C^2	SECV	r_{CV}^2	<i>n</i>	Mean (%)	SD	SEP	r_V^2	RPD
Oil	112	38.78	18.70–46.90	3.88	5	0.541	0.98	0.606	0.98	36	38.45	4.39	0.505	0.99	7.68
Protein	98	19.99	11.89–27.73	3.39	3	1.219	0.88	1.287	0.85	33	19.73	3.32	1.228	0.86	2.76
Carbon	91	57.39	51.41–60.69	1.25	3	0.424	0.88	0.460	0.86	30	57.39	1.04	0.432	0.82	2.90
Nitrogen	90	4.34	3.14–5.52	0.43	4	0.081	0.96	0.091	0.95	30	4.32	0.41	0.087	0.96	4.91

n, number of samples; SD, standard deviation; PLS, partial least square; SEC, standard error of calibration; r_C^2 , coefficient of determination in calibration; SECV, standard error of cross-validation; r_{CV}^2 , coefficient of determination in cross-validation; r_V^2 , coefficient of determination in validation; SEP, standard error of prediction; RPD, ratio of performance deviation (SD/SEP).



for protein, explaining the most important part of the total phenotypic variation except for protein. However, culture effect explained an important part of the total phenotypic variation

varying from 19 to 39% for oil and protein respectively. This result showed that nitrogen and protein contents were more influenced by the culture than oil and carbon.

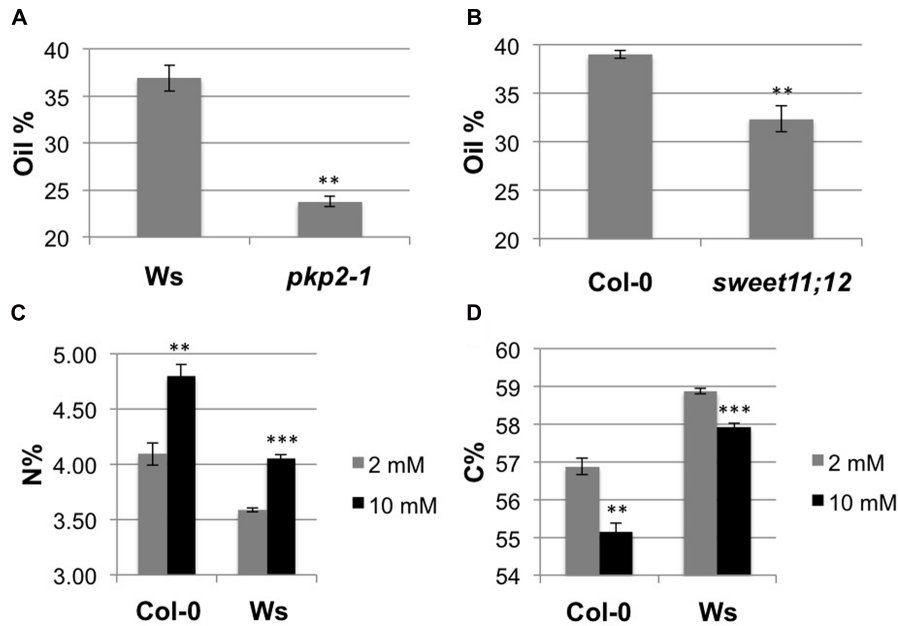


FIGURE 2 | Near-infrared reflectance spectroscopic phenotyping allows detecting seed filling modifications. (A,B) Graphs showing seed oil content of two mutants: *pkp2-1* (A) and *sweet11;12* (B) compared to wild-type. (C,D) Graphs showing seed nitrogen (C) and carbon (D) content of two accessions (Col-0 and Ws) on two nitrogen nutrition conditions (2 and 10 mM). Bars represent SE ($n = 3$ in A,B, $n = 4$ in C,D). Significance in *t*-test, ** $p < 0.01$, *** $p < 0.001$.

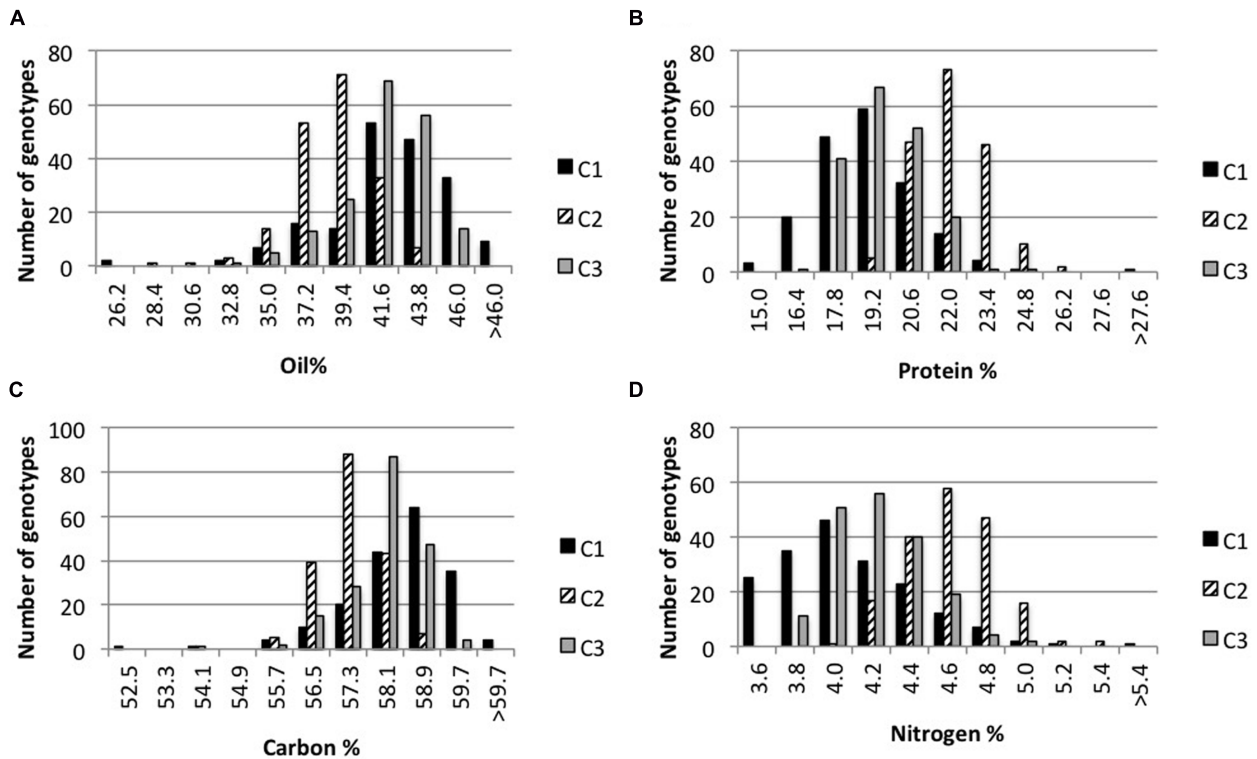


FIGURE 3 | Frequency distribution of seed oil, protein, carbon, and nitrogen content in three independent cultures. Frequency distributions of oil (A), protein (B), carbon (C), and nitrogen (D) content mean values for 183 genotypes ($n = 3$) corresponding to *Arabidopsis* accessions and RILs from eight populations. The same genotypes were cultivated in three independent cultures (C1, C2, C3).

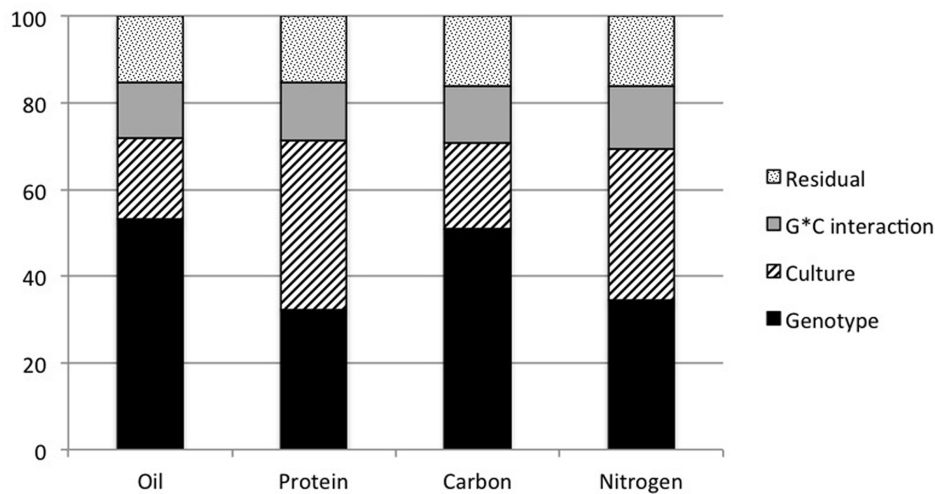


FIGURE 4 | Variance component analysis for seed oil, protein, carbon, and nitrogen content. An ANOVA was performed on all genotypes from the three cultures for seed oil, protein, carbon, and nitrogen content. Histograms show the effects due to genotype, culture, interaction genotype* culture and the residual as a percentage of the variation explained.

Identification of QTL Involved in Seed Oil and Protein Content

The availability of NIRS predictive equations together with the large range of variation observed and the high contribution of genetic part to phenotypic variation for oil, protein, carbon, and nitrogen content opened the door to QTL study. From this previous study of natural variation, the Ct-1 × Col-0 RIL population was selected for QTL determination. A subset of 164 RILs, optimized for QTL mapping (Simon et al., 2008) was cultivated (see Methods) and seeds were phenotyped for oil, protein, carbon, and nitrogen content by NIRS. These RILs exhibited a wide range of values for these four traits as well as transgression beyond the parental line values (not shown), highlighting the potential of this subset to study the variation of these traits. QTL detection using standard procedures (see Methods) was carried out, allowing QTL detection for oil and protein content (this work) as well as for carbon, and nitrogen content (Chardon et al., 2014).

Five QTL for seed oil content were identified, explaining 44.5% of the total phenotypic variance observed (Table 2). The strongest QTL (Oil.4, explaining more than 15% of the phenotypic variance) is located between 31.6 and 46 cM on chromosome 4. Four QTL were detected for protein content, explaining 34% of the total phenotypic variance observed for this trait (Table 2). The strongest QTL (Prot.3, explaining more than 10% of the phenotypic variance) co-localized with Oil.4 on chromosome 4. Three out of five oil QTL, Oil.2, Oil.4, and Oil.5 overlapped with protein QTL Prot.1, Prot.3, and Prot.4 respectively, but having an opposite effect on the corresponding traits, highlighting the strong negative correlation observed between oil and protein seed content, as already observed for carbon and nitrogen by Chardon et al. (2014). Interestingly, QTL specific to oil (Oil.1, i.e., without noticeable effect on protein) and protein (Prot.2, i.e., without noticeable effect on oil) were also identified in this

study. Altogether, these results illustrate that NIRS phenotyping of mature seeds allow identification of genetic factors involved in different pathways of oil and protein accumulation.

DISCUSSION

Quantitative genetic relies on statistical links between phenotype and genotype of 100s of lines. In *Arabidopsis*, genotyping is no more a limiting factor, whereas high-throughput phenotyping can be an obstacle. Thus, our first objective was to establish near-infrared reflectance spectroscopic (NIRS) predictive equations in order to estimate oil, protein, carbon, and nitrogen content in *Arabidopsis* seed with high-throughput level.

TABLE 2 | List of QTL detected for oil and protein in the Ct-1 × Col-0 RIL population.

QTL name	Chr.	Position (cM)	LOD score	CI (cM)	Additive effect	R ² (%)
Oil.1	2	66.9	3.47	47.7–70.8	0.8	4.93
Oil.2	3	6.0	4.57	0–15.8	−0.95	4.73
Oil.3	3	63.6	7.65	48.9–70.9	1.28	9.86
Oil.4	4	36.7	15.30	31.6–46	1.61	15.65
Oil.5	5	13.3	5.58	8.9–17.4	1.82	9.38
Prot.1	3	6.0	4.62	0–11.8	0.58	9.29
Prot.2	4	4.1	4.60	0–12.4	0.61	9.25
Prot.3	4	36.7	5.34	31.6–46	−0.65	10.23
Prot.4	5	13.3	3.84	4.5–17.4	−0.74	5.23

For each QTL, the chromosome (Chr.), the position of the nearest marker of the LOD score peak with the LOD score at the corresponding marker as well as the confidence interval (CI) are indicated. The additive effect represents the mean effect on trait of the replacement of both Col-0 alleles by Ct-1 alleles at the QTL. R² represents the proportion of phenotypic variance of the trait explained by the QTL. cM, centiMorgans.

Near-infrared reflectance spectroscopic calibration models for these four components were established on entire seeds using PLSs regression and leave-one-out cross-validation technique. To assess the accuracy of each model, an external validation with samples not included in the initial model was carried out.

The four developed models display good performances as evaluated by different parameters of the external validation set such as r_V^2 , coefficient of determination; SEP and RPD (Table 1). The r_V^2 values range from 0.82 for carbon content to 0.99 for oil content, indicating that the four models developed in this paper show good to excellent quantitative information (Font et al., 2006). As expected, $SEC_V \geq SEC$ (and then $R_{CV}^2 \leq R_C^2$) for the four models and $SEP \geq SEC$ (and then $R_V^2 \leq R_C^2$) for three out of the four models (Table 1). $SEP < SEC$ (and $R_V^2 > R_C^2$) for the oil model, which is unexpected and illustrates that the restricted validation set (36 samples, i.e., about one third of calibration sample number) fits better to the model. For each model, the validation set shows statistics very close to the calibration set, illustrating robustness and absence of overfitting of the models. Concerning oil content, the model described in this paper for *Arabidopsis* display better performance than the ones described for rapeseed (Tkachuk, 1981; Hom et al., 2007; Rossato et al., 2013). Interestingly, the models were developed on entire seeds without destruction neither any special sample preparation, which is a great advantage compared to calibration developed on powder or oil for example (Khamchum et al., 2013) as it's faster and allow the seeds to be used for other applications.

Seeds produced by a plant are heterogeneous (in size and composition) depending of their position on the mother plant and the environmental conditions during their development. This could induce huge phenotypic variation when phenotyping is performed on very little amount of seeds. The protocol described in this paper overcomes this problem since spectra are determined on a large number of seeds (160 mg, i.e., about 8000 seeds), allowing robust sampling. In favorable environmental condition, an *Arabidopsis* plant produces on average 1 g of seeds, highlighting that the quantity required for NIRS analysis is not a limiting factor. However, in stressful conditions, *Arabidopsis* may produce very few seeds. In this case, NIRS will not be suitable for seed content analysis.

Using *pkp2* and *sweet11;12* described mutants and two different nitrogen nutrition, we demonstrated that NIRS is a powerful method to determine *Arabidopsis* seed composition and that NIRS can probably replace labor intensive methods such as fatty acyl methyl ester extraction followed by Gas chromatography analysis for lipids or elemental analyzer measurements for nitrogen and carbon content.

With NIRS calibrations in hand, natural variation of *Arabidopsis* seed composition was explored. Three independent cultures (C1, C2, and C3) of a 48 core-collection, Col-0 and minimal sets of eight RIL populations were performed, allowing the estimation of environmental effect on the four seed traits analyzed (seed lipid, protein, carbon, and nitrogen content). As shown in Figures 3 and 4, the four traits display a wide range of variation and are strongly impacted by the environment. However, most of the genotypes (75, 97, and 91% from C1, C2, and C3 respectively) display oil content between 32.8 and

43.8%, as already observed by O'Neill et al. (2003) while studying 360 accessions. Similarly, in our three experiments, Cvi-0 was recorded with low oil content (36.21, 33.56, and 34.90%) while Ct-1 was recorded with high oil content (46.26, 42.40, and 45.3%) as in O'Neill et al. (2003). Even though seed composition is strongly impacted by environmental conditions, the four traits analyzed are also controlled by genetic factor as illustrated by Figure 4. Indeed nine QTL were identified for seed oil and protein content in the Ct-1 × Col-0 RIL population (Table 2). Most of the QTL for oil content co-localized with QTL for protein content but with opposite effect on each traits, highlighting the negative correlation between seed oil and protein content. Oil.1 and Oil.2 co-localized with seed oil content QTL previously identified by Hobbs et al. (2004) in the Ler × Cvi-0 RIL population and by O'Neill et al. (2012) in the Cvi-0 × Ag-0 RIL population for Oil.1. Interestingly, Oil.1 does not co-localize with seed protein QTL in the Ct-1 × Col-0 RIL population, suggesting that Oil.1 may regulate oil content without affecting protein content and thus represents a very good candidate to specifically modify oil content without affecting protein content in *Arabidopsis* seed. Conversely, Prot.2 may regulate protein content independently of oil content and could be used to solely modify seed protein content. Fine mapping is required to confirm Oil.1 and Prot.2 specificity as well as to identify the genes under the nine QTL identified.

CONCLUSION

In summary, the results of the present work show that NIRS predictive equations developed in this study can be used to reliably predict oil, protein, nitrogen and carbon content of *Arabidopsis* seed samples without destruction neither any special sample preparation. This high-throughput method opens the way for quantitative genetic such as QTL cloning (up to gene identification and not only detection), as well as GWASs but also to mutant library screening. As a first attempt to identify genetic factor controlling seed oil and protein content, QTL for these traits have been mapped in the *Arabidopsis* Ct-1 × Col-0 RIL population. Some of the oil content QTL detected co-localized with QTL identified previously, thus validated our approach, but many novel QTL were also identified. In particular, to our knowledge, this is the first report of seed protein content QTL in *Arabidopsis*. The fine mapping of some of these QTL is underway and should give new insights on the regulatory pathway involved in *Arabidopsis* seed oil and protein accumulation.

METHODS

Plant Material

The 48 core-collection of *Arabidopsis* (McKhann et al., 2004) in addition to Col-0 accession, *wri1-3*, *wri1-4*, *tag1-2*, *pkp2-1*, and *sweet11;12* mutants, a minimal set (20 lines) of eight RIL populations (2RV, 3RV, 7RV, 8RV, 13RV, 17RV, 20RV, and 21RV) as well as the core-pop of 164 RILs of the Ct-1 × Col-0

population were used in this study (Simon et al., 2008). *wri1-3*, *wri1-4*, *pkp2-1*, and *tag1-2* seeds were provided by S. Baud and *sweet11;12* seeds were provided by R. Le Hir. The other seeds were obtained from the Versailles Biological Resource Centre for *Arabidopsis*¹. Seeds were sown on damp Whatman filters, stratified for 3 days at 4°C and then transferred to a growth cabinet under long-day conditions at 21°C for 2 days. Three seedlings (with emerging radicle) per genotype were planted in soil in 7 cm pots and transferred to a non-heated and naturally lit greenhouse to be vernalized from November to February. After 8 weeks, one plantlet per pot was randomly retained without phenotype selection. After 12 weeks of vernalization, plants were transferred to a growth chamber under long-day conditions (16/8 h photoperiod at 150 mmol photons m⁻² s⁻¹); 21°C day temperature and 18°C night temperature; relative humidity of 65%. From this time, three times a week the plant trays were moved around the growth chamber to reduce position effects. Bags were put over the plants to prevent seed dispersion as soon as the first silique had turned yellow. The plants were no longer watered once the youngest silique had turned yellow. Plants were kept in the growth chamber until dry and then harvested.

Three cultures (C1, C2, and C3) including the 48 core-collection, Col-0, the minimal sets of the 8 RIL populations and *wri1-3*, *wri1-4*, and *tag1-2* mutants for C3, as well as one culture (C4) of the core-pop of 164 RILs of the Ct-1 × Col-0 population were performed following this protocol.

Near-Infrared Spectroscopy NIR Spectra Acquisition

Seed samples were placed in a 9 mm diameter clear glass bottle (Agilent, 5182-0714) on 4 mm height for NIRS spectra acquisition and were analyzed as intact (without any treatment). This corresponds to about 300 μl of *Arabidopsis* seeds (about 160 mg or 8 000 seeds).

Spectra acquisition was performed with a Fourier transform near-infrared (FTNIR) analyzer (Antaris II spectrometer; Thermofisher Scientific, France). Spectra were collected in reflectance mode with an 8 cm⁻¹ optical resolution and were obtained as an average of 16 scans. Spectra were collected over the range 4000 to 10000⁻¹ and calibrations done using four spectral ranges: from 4100 to 4940 cm⁻¹; from 5390 to 6690 cm⁻¹; from 6900 to 7130 cm⁻¹, and from 7185 to 9000 cm⁻¹. These spectral regions provide useful information about the organic signature of the *Arabidopsis* samples and exclude the water spectral regions. They have been selected by looking at the regression vector from the PLS (see Development of NIRS Calibration Models) and using a Thermo proprietary pure component algorithm.

Selecting the Samples for NIRS Calibration

The robustness and accuracy of a NIRS model are strongly dependent on the accuracy of the reference method but also of the samples chosen for calibration development. Indeed, the calibration samples have to be representative of the

spectral variability and must cover the range of the component concentration of the samples that will be further monitored.

As the NIR spectral variability of *Arabidopsis* seeds was not known, NIR spectra of 650 samples (one spectrum per sample) from two independent cultures (C1 and C2) were collected. Spectra were treated with a multiplicative signal correction (MSC) to correct multiplicative effects due to light scattering in spectral data and a Principal component analysis (PCA) was performed in order to select samples maximizing spectral variability. PC1 and PC2 explained 84.5 and 11.9% of the spectral variation respectively. Their graphic representations were similar to a seed spectrum and suggested that they reflect variations due to differences in spectra baseline or particle size for example. PC3 explained 1.9% of the spectral variation and its graphic representation displayed peaks at wavelengths specific to seed storage compounds. Thus the PC1/PC3 graph was used as a criterion for selecting 100 samples in the population as being more variable on the basis of spectra features (Shenk and Westerhaus, 1991).

Seed oil content of these 100 samples was determined by the Soxhlet reference method and a preliminary calibration model was set up. Using this model, seed oil content of samples available at this time (1788 samples from three independent cultures and including *wri1* and *tag1* low-seed-oil insertion mutants) was predicted. This prediction allowed the selection of 48 additional samples with extreme values (maximal and minimal) in order to extend the range of concentration of the final calibration set. Seed oil content was measured with the Soxhlet method on these 48 additional samples. The same procedure was applied to choose samples for seed protein, carbon, and nitrogen content calibration models.

Development of NIRS Calibration Models

Calibration models were developed using TQ Analyst software (Thermofisher Scientific, France) using PLSs regression and leave-one-out cross-validation technique. Prior to the PLS regression, all spectra were pre-treated with the scatter correction MSC and by applying a first derivative transformation and a Norris derivative filter (segment length 5, gap size: 5). The use of derivative spectra instead of the raw optical data to perform calibration is a way of solving problems associated with offsets and overlapping peaks.

Near-infrared reflectance spectroscopic calibration models were established for oil, protein, carbon, and nitrogen content by using a number of PLS factors optimal for each component (i.e., only the primary, most important factors were used, the “noise” being encapsulated in the less important factors). The optimal number of PLS factors was determined as the minimum of the PRESS (predicted residual error sum of squares) curve when doing a leave-one-out cross validation method.

The quality of each calibration model was then evaluated by several parameters: the determination coefficients between concentrations predicted from NIRS and from reference analysis, r_c^2 and r_{CV}^2 , calculated for calibration and cross-validation (leave-one-out) data processing respectively, and their respective standard errors [calibration (SEC) and cross-validation (SECV)].

¹<http://publiclines.versailles.inra.fr/>

To assess the accuracy of each newly developed calibration model, an external validation with samples not included in the initial model was carried out. The total number of samples was divided into calibration and external validation sets in a rate 3:1. For that purpose, the samples were ranked according to their reference values and then about one sample every four was assigned to the external validation set. In addition, to account for environmental variation in seed composition, the seed samples chosen for the calibration and external validation sets were derived from the three cultures (C1, C2, and C3). The prediction quality of NIRS analyses was then quantified by the SEP and the determination coefficient (r^2) between concentrations obtained from NIRS and from reference analysis for the validation set. The RPD was calculated as the ratio between the SD of the reference values and the SEP. RPD is indicative of the usefulness of the NIRS calibrations.

Analysis of Seed Oil Content (Reference Method)

Oil was extracted following the standard NF V03-908 protocol (extraction by hot solvent with a “Soxhlet” extractor). About 1 g of seed was dried (103°C during 20 h) and ground in hexane with a grinder. Oil was then extracted with hexane by the Soxhlet method. The total seed oil content was expressed as percentage of the dried seed weight.

Analysis of Seed Protein Content (Reference Method)

Phenol extraction of seed protein was adapted from Meyer et al. (1988). Ten mg of seeds were homogenized in a 2 ml tube containing a ceramic bead and 1 ml of an emulsion of 50% (v/v) phenol (previously equilibrated in 1 M Tris HCl pH8) in 0.1 M Tris HCl pH8 1% SDS using a Fastprep-24 Instrument (MP-Biomedical, maximal intensity, twice 1 min). After centrifugation (13 000 g, 20 min), 200 μ l of the phenolic phase was accurately delipidated twice with 500 μ l of hexane. One hundred μ l of the phenol phase was taken after centrifugation (13 000 g, 10 min) and the proteins were precipitated with five volumes of methanol containing 0.1 M ammonium acetate at -20°C overnight. The precipitate was collected by centrifugation and washed four times with methanol (-20°C) containing 0.1 M ammonium acetate, and twice with 80% acetone (in water). The resulting pellets were dried under reduced pressure and then resuspended in 1 ml of 0.1 M Tris-HCl pH8 1% SDS. After overnight agitation, the fully dissolved solution was then cleared by centrifugation (13 000 g, 10 min) and the protein concentration was determined by spectrometry at 280 nm, assuming that 1 OD corresponds to 1 mg/ml protein solution.

Seed Nitrogen and Carbon Content (Reference Method)

Five mg of seeds, dried overnight at 100°C , were weighed on a lab balance model M2P (Sartorius, Göttingen, Germany) with a readability of 0.001 mg, then analyzed for nitrogen and carbon concentration by the Dumas combustion method (Anonymous,

1990) with an automated CN analyzer (Heraeus CN-Rapid, Hanau, Germany).

QTL Detection

For each RIL, the mean value from three plants was taken for each measured trait for QTL analysis.

Quantitative trait loci analyses were performed using R/qtl library in the R environment (Broman et al., 2003; Arends et al., 2010) with standard methods for interval mapping (IM) and multiple QTL mapping (MQM) (Arends et al., 2010). First, IM was carried out to determine putative QTL involved in the variation of the trait, and then MQM model was performed on the same data: the closest marker to each local logarithm-of-odds (LOD) score peak (putative QTL) was used as a cofactor to control the genetic background while testing at another genomic position. The significance threshold ($p < 0.05$) of LOD was determined by permutation test ($n = 1000$) for each trait (Churchill and Doerge, 1994). The estimated additive effect (representing the mean effect of the replacement of the *Col-0* alleles by *Ct-1* alleles at the locus) and the percentage of variance explained by each QTL (R^2) affecting a trait were obtained for the final MQM model.

AUTHOR CONTRIBUTIONS

PG, PB-M, and SJ established NIRS models. PG and SJ performed the statistical analysis and wrote the manuscript. AL and MD carried out the plant cultures, seed protein content (reference method) measurements and NIR spectrum acquisitions and predictions. PG and AL performed seed carbon and nitrogen content analysis (reference method). SJ performed the QTL detection experiments. All authors read and approved the final manuscript.

FUNDING

This work was supported by the INRA “Biology and Plant Breeding” Department. The IJPB benefits from the support of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS).

ACKNOWLEDGMENTS

Soxhlet oil extractions were performed by M. Krouti (Terres Inovia, Ardon, France). We thank F. Chardon and M. Reymond for critical reading of the manuscript. We thank S. Baud and R. Le Hir for *pkp2-1* and *sweet11;12* seeds respectively, as well as A. Marmagne for *Col-0* and *Ws* seeds from plants grown on 2 and 10 mM nitrate. We also thank Lilian Dahuron, Philippe Marechal, and Sébastien Bénard for plant care.

REFERENCES

- AACC International (1999). *Approved Methods of Analysis, Method 39-00.01. Near-Infrared Methods—Guidelines for Model Development and Maintenance. Approved November 3, 1999*, 11th Edn. St. Paul, MN: AACC International. doi: 10.1094/AACCIntMethod-39-00.01
- Anonymous (1990). "Crude protein in animal feed combustion method," in *Official Methods of Analysis*, 15th Edn, ed. K. Helrich (Rockville, MD: Association of Official Analytical Chemists, Inc.).
- Arends, D., Prins, P., Jansen, R. C., and Broman, K. W. (2010). R/qtl: high-throughput multiple QTL mapping. *Bioinformatics* 26, 2990–2992. doi: 10.1093/bioinformatics/btq565
- Batten, G. D. (1998). Plant analysis using near infrared reflectance spectroscopy: the potential and the limitations. *Aust. J. Exp. Agric.* 38, 697–706. doi: 10.1071/ea97146
- Baud, S., Wullemme, S., Dubreucq, B., de Almeida, A., Vuagnat, C., Lepiniec, L., et al. (2007). Function of plastidial pyruvate kinases in seeds of *Arabidopsis thaliana*. *Plant J.* 52(3), 405–419. doi: 10.1111/j.1365-313X.2007.03232.x
- Beisson, F., Koo, A. J., Ruuska, S., Schwender, J., Pollard, M., Thelen, J. J., et al. (2003). *Arabidopsis* genes involved in acyl lipid metabolism. A 2003 census of the candidates, a study of the distribution of expressed sequence tags in organs, and a web-based database. *Plant Physiol.* 132, 681–697. doi: 10.1104/pp.103.022988
- Branham, S. E., Wright, S. J., Reba, A., and Linder, C. R. (2015). Genome-wide association study of *Arabidopsis thaliana* identifies determinants of natural variation in seed oil composition. *J. Hered.* 107, 248–256. doi: 10.1093/jhered/esv100
- Broman, K. W., Wu, H., Sen, S., and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* 19, 889–890. doi: 10.1093/bioinformatics/btg112
- Chardon, F., Jasinski, S., Durand, M., Lecureuil, A., Soulay, F., Bedu, M., et al. (2014). QTL meta-analysis in *Arabidopsis* reveals an interaction between leaf senescence and resource allocation to seeds. *J. Exp. Bot.* 65, 3949–3962. doi: 10.1093/jxb/eru125
- Chen, L. Q., Lin, I. W., Qu, X. Q., Sosso, D., McFarlane, H. E., Londono, A., et al. (2015). A cascade of sequentially expressed sucrose transporters in the seed coat and endosperm provides nutrition for the *Arabidopsis* embryo. *Plant Cell* 27, 607–619. doi: 10.1105/tpc.114.134585
- Churchill, G. A., and Doerge, R. W. (1994). Empirical threshold values for quantitative trait mapping. *Genetics* 138, 963–971.
- Durrett, T. P., Benning, C., and Ohlrogge, J. (2008). Plant triacylglycerols as feedstocks for the production of biofuels. *Plant J.* 54, 593–607. doi: 10.1111/j.1365-313X.2008.03442.x
- Dyer, J. M., Szymme, S., Green, A. G., and Carlsson, A. S. (2008). High-value oils from plants. *Plant J.* 54, 640–655. doi: 10.1111/j.1365-313X.2008.03430.x
- Eskandari, M., Cober, E. R., and Rajcan, I. (2012). Genetic control of soybean seed oil: I. QTL and genes associated with seed oil concentration in RIL populations derived from crossing moderately high-oil parents. *Theor. Appl. Genet.* 126, 483–495. doi: 10.1007/s00122-012-1995-3
- Focks, N., and Benning, C. (1998). wrinkled1: a novel, low-seed-oil mutant of *Arabidopsis* with a deficiency in the seed-specific regulation of carbohydrate metabolism. *Plant Physiol.* 118, 91–101. doi: 10.1104/pp.118.1.91
- Font, R., del Rio, M., Fernandez, J. M., and de Haro, A. (2003). Acid detergent fiber analysis in oilseed Brassicas by near-infrared spectroscopy. *J. Agric. Food Chem.* 51, 2917–2922. doi: 10.1021/jf021011
- Font, R., del Rio, M., Fernandez-Martinez, J. M., and de Haro-Bailon, A. (2004). Use of near-infrared spectroscopy for screening the individual and total glucosinolate contents in Indian mustard seed (*Brassica juncea* L. Czern. & Coss.). *J. Agric. Food Chem.* 52, 3563–3569. doi: 10.1021/jf0307649
- Font, R., del Rio-Celestino, M., and de Haro-Bailon, A. (2006). The use of near-infrared spectroscopy (NIRS) in the study of seed quality components in plant breeding programs. *Ind. Crops Prod.* 24, 307–313. doi: 10.1016/j.indcrop.2006.06.012
- Font, R., Rio, M. D., Fernandez, J., Arthur, E., Bancroft, I., Chinoy, C., et al. (2002a). Seed oil content analysis of Ethiopian mustard (*Brassica carinata* A. Braun) by near infrared spectroscopy. *Cruciferae Newsletter* 24, 5–6.
- Font, R., Rio, M. D., Sillero, A., Arthur, E., Bancroft, I., Chinoy, C., et al. (2002b). Using near infrared spectroscopy for determining protein content in Ethiopian mustard (*Brassica carinata* A. Braun). *Cruciferae Newsletter* 24, 7–8.
- Hobbs, D. H., Flintham, J. E., and Hills, M. J. (2004). Genetic control of storage oil synthesis in seeds of *Arabidopsis*. *Plant Physiol.* 136, 3341–3349. doi: 10.1104/pp.104.049486
- Hom, N. H., Becker, H. C., and Mollers, C. (2007). Non-destructive analysis of rapeseed quality by NIRS of small seed samples and single seeds. *Euphytica* 153, 27–34. doi: 10.1007/s10681-006-9195-3
- Hwang, E. Y., Song, Q., Jia, G., Specht, J. E., Hyten, D. L., Costa, J., et al. (2014). A genome-wide association study of seed protein and oil content in soybean. *BMC Genomics* 15:1. doi: 10.1186/1471-2164-15-1
- Irzykowska, L., and Wolko, B. (2004). Interval mapping of QTLs controlling yield-related traits and seed protein content in *Pisum sativum*. *J. Appl. Genet.* 45, 297–306.
- Jasinski, S., Lecureuil, A., Miquel, M., Loudet, O., Raffaele, S., Froissard, M., et al. (2012). Natural variation in seed very long chain fatty acid content is controlled by a new isoform of KCS18 in *Arabidopsis thaliana*. *PLoS ONE* 7:e49261. doi: 10.1371/journal.pone.0049261
- Khamchum, C., Punsuvon, V., Kasemsumran, S., and Suttiwijitpukdee, N. (2013). A feasibility study of oil content and fatty acid composition of seed powder and seed oil of *Pongamia pinnata* by near infrared spectroscopy. *Scienceasia* 39, 384–391. doi: 10.2306/scienceasia1513-1874.2013.39.384
- Kianian, S. F., Egli, M. A., Phillips, R. L., Rines, H. W., Somers, D. A., Gengenbach, B. G., et al. (1999). Association of a major groat oil content QTL and an acetyl-CoA carboxylase gene in oat. *Theor. Appl. Genet.* 98, 884–894. doi: 10.1007/s001220051147
- Kumar, S., You, F. M., Duguid, S., Booker, H., Rowland, G., and Cloutier, S. (2015). QTL for fatty acid composition and yield in linseed (*Linum usitatissimum* L.). *Theor. Appl. Genet.* 128, 965–984. doi: 10.1007/s00122-015-3
- Li, F., Chen, B., Xu, K., Wu, J., Song, W., Bancroft, I., et al. (2014). Genome-wide association study dissects the genetic architecture of seed weight and seed quality in rapeseed (*Brassica napus* L.). *DNA Res.* 21, 355–367. doi: 10.1093/dnares/dsu002
- Li, H., Peng, Z., Yang, X., Wang, W., Fu, J., Wang, J., et al. (2012). Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat. Genet.* 45, 43–50. doi: 10.1038/ng.2484
- Liu, G. Z., Mei, H. X., Wang, S., Li, X. H., Zhu, X. F., and Zhang, T. Z. (2015). Association mapping of seed oil and protein contents in upland cotton. *Euphytica* 205, 637–645. doi: 10.1007/s10681-015-1450-z
- Liu, P., Wang, C. M., Li, L., Sun, F., and Yue, G. H. (2011). Mapping QTLs for oil traits and eQTLs for oleosin genes in jatropha. *BMC Plant Biol.* 11:132. doi: 10.1186/1471-2229-11-132
- Masclaux-Daubresse, C., and Chardon, F. (2011). Exploring nitrogen remobilization for seed filling using natural variation in *Arabidopsis thaliana*. *J. Exp. Bot.* 62, 2131–2142. doi: 10.1093/jxb/erq405
- McKhann, H. I., Camilleri, C., Berard, A., Bataillon, T., David, J. L., Reboud, X., et al. (2004). Nested core collections maximizing genetic diversity in *Arabidopsis thaliana*. *Plant J.* 38, 193–202. doi: 10.1111/j.1365-313X.2004.02034.x
- Meyer, Y., Grosset, J., Chartier, Y., and Cleyetmarel, J. C. (1988). Preparation by two-dimensional electrophoresis of proteins for antibody-production - antibodies against proteins whose synthesis is reduced by auxin in tobacco mesophyll protoplasts. *Electrophoresis* 9, 704–712. doi: 10.1002/elps.1150091105
- Mokrani, L., Gentzittel, L., Azanza, F., Fitamant, L., Al-Chaarani, G., and Sarrafi, A. (2002). Mapping and analysis of quantitative trait loci for grain oil content and agronomic traits using AFLP and SSR in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* 106, 149–156.
- O'Neill, C. M., Gill, S., Hobbs, D., Morgan, C., and Bancroft, I. (2003). Natural variation for seed oil composition in *Arabidopsis thaliana*. *Phytochemistry* 64, 1077–1090. doi: 10.1016/S0031-9422(03)00351-0
- O'Neill, C. M., Morgan, C., Hattori, C., Brennan, M., Rosas, U., Tsochoep, H., et al. (2012). Towards the genetic architecture of seed lipid biosynthesis and accumulation in *Arabidopsis thaliana*. *Heredity* 108, 115–123. doi: 10.1038/hdy.2011.54
- Parkin, I. A., Gulden, S. M., Sharpe, A. G., Lukens, L., Trick, M., Osborn, T. C., et al. (2005). Segmental structure of the *Brassica napus* genome based on

- comparative analysis with *Arabidopsis thaliana*. *Genetics* 171(2), 765–781. doi: 10.1534/genetics.105.042093
- Plessis, A., Ravel, C., Bordes, J., Balfourier, F., and Martre, P. (2013). Association study of wheat grain protein composition reveals that gliadin and glutenin composition are trans-regulated by different chromosome regions. *J. Exp. Bot.* 64, 3627–3644. doi: 10.1093/jxb/ert188
- Rossato, R., Prete, C. E. C., de Castro, C., Tomm, G. O., Leite, R. S., Mandarino, J. M. G., et al. (2013). Predicting rapeseed oil content with near-infrared spectroscopy. *Pesqui. Agropecuaria Bras.* 48, 1601–1604. doi: 10.1590/s0100-204x2013001200010
- Sanyal, A., and Linder, C. R. (2012). Quantitative trait loci involved in regulating seed oil composition in *Arabidopsis thaliana* and their evolutionary implications. *Theor. Appl. Genet.* 124, 723–738. doi: 10.1007/s00122-011-1742-1
- Shenk, J. S., and Westerhaus, M. O. (1991). Population structuring of near-infrared spectra and modified partial least-squares regression. *Crop Sci.* 31, 1548–1555. doi: 10.2135/cropsci1991.0011183X003100060034x
- Shewry, P. R., Napier, J. A., and Tatham, A. S. (1995). Seed storage proteins: structures and biosynthesis. *Plant Cell* 7, 945–956. doi: 10.1105/tpc.7.7.945
- Simon, M., Loudet, O., Durand, S., Berard, A., Brunel, D., Sennesal, F. X., et al. (2008). Quantitative trait loci mapping in five new large recombinant inbred line populations of *Arabidopsis thaliana* genotyped with consensus single-nucleotide polymorphism markers. *Genetics* 178, 2253–2264. doi: 10.1534/genetics.107.083899
- Sun, M., Hua, W., Liu, J., Huang, S., Wang, X., Liu, G., et al. (2012). Design of new genome- and gene-sourced primers and identification of QTL for seed oil content in a specially high-oil *Brassica napus* cultivar. *PLoS ONE* 7:e47037. doi: 10.1371/journal.pone.0047037
- Tar'an, B., Warkentin, T., Somers, D. J., Miranda, D., Vandenberg, A., Blade, S., et al. (2004). Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L.). *Euphytica* 136, 297–306. doi: 10.1023/B:EUPH.0000032721.03075.a0
- Tkachuk, R. (1981). Oil and protein-analysis of whole rapeseed kernels by near-infrared reflectance spectroscopy. *J. Am. Oil Chem. Soc.* 58, 819–822. doi: 10.1007/bf02665588
- Valentini, G., Chiarelli, L., Fortin, R., Speranza, M. L., Galizzi, A., and Mattevi, A. (2000). The allosteric regulation of pyruvate kinase. *J. Biol. Chem.* 275, 18145–18152. doi: 10.1074/jbc.M001870200
- Velasco, L., and Becker, H. C. (1998). Analysis of total glucosinolate content and individual glucosinolates in *Brassica* spp. by near-infrared reflectance spectroscopy. *Plant Breed.* 117, 97–102. doi: 10.1111/j.1439-0523.1998.tb01459.x
- Yang, X., Ma, H., Zhang, P., Yan, J., Guo, Y., Song, T., et al. (2012). Characterization of QTL for oil content in maize kernel. *Theor. Appl. Genet.* 125, 1169–1179. doi: 10.1007/s00122-012-1903-x
- Ying, J. Z., Shan, J. X., Gao, J. P., Zhu, M. Z., Shi, M., and Lin, H. X. (2012). Identification of quantitative trait loci for lipid metabolism in rice seeds. *Mol. Plant* 5, 865–875. doi: 10.1093/mp/ssp100
- Zheng, P., Allen, W. B., Roesler, K., Williams, M. E., Zhang, S., Li, J., et al. (2008). A phenylalanine in DGAT is a key determinant of oil content and composition in maize. *Nat. Genet.* 40, 367–372. doi: 10.1038/ng.85

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Jasinski, Lécureuil, Durandet, Bernard-Moulin and Guerche. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.